Asian TeX-like typeset engines

Jean-Michel Hufflen

Asian TEX-Like Typeset Engines*

Jean-Michel HUFFLEN
LIFC (EA CNRS 4157)
University of Franche-Comté
16, route de Gray
25030 BESANÇON CEDEX
FRANCE
hufflen@lifc.univ-fcomte.fr
http://lifc.univ-fcomte.fr/~hufflen

Abstract

In order to extend MlBibTEXto languages of the Far East, we are experiencing TEX engines for them — e.g., pTEX — after attending the first Asian TEX conference. We propose a demonstration of that.

Keywords Asian TEX engines, pTEX, CJK package, kotex package.

Streszczenie

Przy rozszerzaniu MlBibTEX-a do pracy z językami Dalekiego Wschodu, dzięki udziałowi w 1-szej konferencji Asia TEX, mogliśmy zająć się odpowiednimi silnikami TEX-owymi, np. pTEX-em. Zademostrujemy wyniki.

Słowa kluczowe warianty TEXa dla języków dalekowschodnich, pTEX, pakiet CJK, pakiet kotex.

0 Introduction

In January 2008, the first Asian TEX conference was held at Kongju, in South Korea. In fact, presentations focused on languages used throughout the Far East, since most people attending at this conference came from Korea, China, Japan, and Vietnam. Concerning us, we are interested in enlarging MlBibTEX[1] [5], our reimplementation of BibTEX [14], to Asian languages. Going to this conference allowed us to get more familiar with the packages and engines used for these languages, for which only a little documentation is written in English.

Roughly speaking, the present situation about Asian languages is the same than at the beginning of the 1990's for European languages: some *ad hoc* packages — french [3], german [16], polish [2, § F.7] — were already being developed, but without global organisation. Each package allowed users to write both in their own language and English. The babel package was only announced in *The LATEX Companion*'s first edition [4, Ch. 9]; it is now fully working and described in *The LATEX Companion*'s second edition [12, Ch. 9]. Let us recall that this package aims to process all the languages it knows in a homogeneous way[2]. However, the present distribution does not include support for Far East languages by means of this package, although some tools presented hereafter — the packages CJK and kotex — can be used in conjunction with it.

Section 1 gives a short overview of the CJK package. Then we cite two *ad hoc* ways to write in Korean and Japanese in Sections 2 and 3. In addition to these packages and engines, let us just cite VnTEX for the Vietnamese language, including Vietnamese fonts[3], support for LATEX and Plain TEX [18]. The definitions suitable for LATEX are usable by means of a vietnam package or a vietnam option of the babel package.

1 The CJK package

'CJK' is a collective term for '**C**hinese, **J**apanese, **K**orean', since these writing systems completely or partly use Chinese characters. 'CJK' characters are implemented in an area of Unicode's basic multilingual plane [19], inside the range U+3000–U+9FA5.

* Title in Polish: *TEX-owe narzędzia składu dla języków Dalekiego Wschodu*

[1] **M**ulti**L**ingual BibTEX. This program has been used to build the current article's bibliography. Asian person names are put down according to their tradition, that is, the last name comes before the first name.

[2] In fact, it encompasses most of the features provided by packages such as french, german, polish.

[3] Vietnamese is written with Latin letters, with many accented letters. Some letters can have two accents, the total number of accented letters in Vietnamese being 134 [18].

In fact, the Chinese and Japanese languages use ideograms, whereas modern Korean is written using the Hangul alphabet, organised into syllabic blocks. Each block is a combination of consonants and vowels and notes a syllable down. Traditionally, Chinese and Korean are written horizontally. Often the Japanese language uses vertical writing according to its its tradition, but the horizontal sense is used for technical writing.

The CJK package [10] implements several encoding schemes to reach the characters belonging to the CJK area and some basic typographic rules related to these three languages. It can be used in conjunction with the babel package. Its limitation: that is preferable for it if a non-CJK language is the main language[4]. This package has been improved into a CJKutf8 package [11], dealing with UTF-8[5] texts more easily, and a version derived from it — the xCJK package — is usable with the XeTeX typeset engine[6] [6].

2 The koTeX package

The steps of developing a Hangul TeX are related in Korean in [1]. It seems that Korean people presently use the tools of the koTeX distribution [8], including the kotex package and the halpha bibliography style, usable with BibTeX. This kotex package is able to deal with UTF-8 texts. Text fragments — including texts put within commands' arguments — are supposed to be in Korean (resp. English) if Asian (resp. Latin) characters are used. Besides, it can be used in conjunction with the babel package.

The LaTeX source text of a very short example is given at Figure 1. This complete example — as the ajt-sample.tex file — belongs to the document class files for the *Asian Journal of TeX*[7], we have just replaced the fragments using Korean characters by occurrences of '...'. When this text is processed by LaTeX, the korean option of the ajt document class causes the kotex package to be loaded, and the result is given at Figure 2.

```
%%%  Select pdfTeX or your favorite DVI driver.
\documentclass[korean]{ajt} % pdfTeX (default).
%%%

%%% Mandatory article metadata.
\title[How to prepare a document with AJT
class]{AJT ...}
\author[Gildong Hong]{...}
\address{...}
\email{gdhong@ktug.or.kr}

\abstract{{}... AJT ...(abstract)...
preamble... AJT ... \PracTeX...}
%%%

\begin{document}

\maketitle
%%%

\section{...}
...

\subsection{...}
...

\subsection{...}
... \verb+\[+ ... \verb+\]+ ...

\begin{thebibliography}{9}
\bibitem[1]{...} ...

\bibitem[2]{...} ...

\bibitem[4]{...} ...

\bibitem[5]{...} ...
\end{thebibliography}
%%%

\end{document}
```

Figure 1: Example using the kotex package: source text.

3 Japanese TeX

pTeX [13] is a typeset engine based on TeX and implementing formatting rules for Japanese documents. In particular, it deals with the two possible senses of writing — horizontal or vertical —, provides some *ad hoc* document classes jsarticle, jsbook, jreport. Its new version, upTeX [17] may be viewed as 'a new pTeX', dealing with Unicode and supporting the babel package. An integration of upTeX and Ω[8] [15], upOmega, is also planned.

To end, let us mention the development of a Japanese TeX environment for Cygwin, usable under Windows [9].

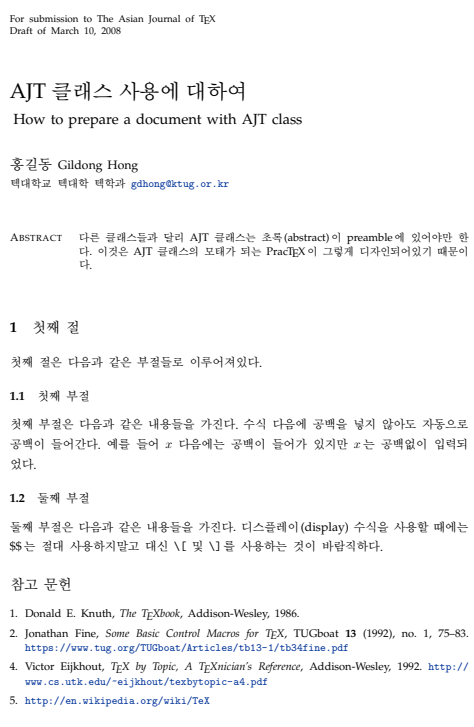
[4] As abovementioned, this CJK package mainly aims to access CJK characters and assemble them into fragments. But — as an example — for the whole of a text written using one of CJK languages, the spacing between two baselines should be wider, that is, the value of the `\baselineskip` command [7, Ch. 12] should be enlarged. So do the kotex package (cf. § 2) and the pTeX engine (cf. § 3).

[5] Unicode **T**ransformation **F**ormat. The UTF-8 encoding is probably the most used now for texts including non-Latin characters. See [19] for more details.

[6] XeTeX is a typesetting system based on a merger of TeX's system with Unicode and modern font technologies.

[7] You can download all these files from `http://ftp.ktug.or.kr/pub/ktug/ajt/`.

[8] Ω is an extension of TeX aiming to improve multilingual typesetting.



For submission to The Asian Journal of TEX
Draft of March 10, 2008

AJT 클래스 사용에 대하여

How to prepare a document with AJT class

홍길동 Gildong Hong
택대학교 택대학 택학과 gdhong@ktug.or.kr

ABSTRACT 다른 클래스들과 달리 AJT 클래스는 초록(abstract)이 preamble에 있어야만 한다. 이것은 AJT 클래스의 모태가 되는 PracTEX이 그렇게 디자인되어있기 때문이다.

1 첫째 절

첫째 절은 다음과 같은 부절들로 이루어져있다.

1.1 첫째 부절

첫째 부절은 다음과 같은 내용들을 가진다. 수식 다음에 공백을 넣지 않아도 자동으로 공백이 들어간다. 예를 들어 x 다음에는 공백이 들어가 있지만 x는 공백없이 입력되었다.

1.2 둘째 부절

둘째 부절은 다음과 같은 내용들을 가진다. 디스플레이(display) 수식을 사용할 때에는 $$는 절대 사용하지말고 대신 \[및 \] 를 사용하는 것이 바람직하다.

참고 문헌

1. Donald E. Knuth, *The TEXbook*, Addison-Wesley, 1986.
2. Jonathan Fine, *Some Basic Control Macros for TEX*, TUGboat **13** (1992), no. 1, 75–83. https://www.tug.org/TUGboat/Articles/tb13-1/tb34fine.pdf
4. Victor Eijkhout, *TEX by Topic, A TEXnician's Reference*, Addison-Wesley, 1992. http://www.cs.utk.edu/~eijkhout/texbytopic-a4.pdf
5. http://en.wikipedia.org/wiki/TeX

Figure 2: Example using the kotex package: result.

4 Acknowledgements

Many thanks to Jerzy B. Ludwichowski, who has written the Polish translation of the abstract. Tomasz Przechlewski has helped translate keywords, thanks to him, too.

References

[1] Cho Jin-Hwan: "The Passage of Hangul TEX and koTEX". *The Asian Journal of TEX*, Vol. 1, no. 2, pp. 113–121. October 2007.

[2] Antoni Diller: *LATEX wiersz po wierszu*. Wydawnictwo Helio, Gliwice. Polish translation of *LATEX Line by Line* with an additional annex by Jan Jelowicki. 2001.

[3] Bernard Gaulle : *Notice d'utilisation du style french multilingue pour LATEX*. Version 4.05b. Février 1999. Joint à la distribution de LATEX.

[4] Michel Goossens, Frank Mittelbach and Alexander Samarin: *The LATEX Companion*. 1st edition. Addison-Wesley Publishing Company, Reading, Massachusetts. 1994.

[5] Jean-Michel Hufflen: "MlBibTEX: Reporting the Experience". TUG*boat*, Vol. 29, no. 1, pp. 157–162. EuroBachoTEX 2007 proceedings. 2007.

[6] Jonathan Kew: "XeTEX in TEX Live and beyond". TUG*boat*, Vol. 29, no. 1, pp. 146–150. EuroBachoTEX 2007 proceedings. 2007.

[7] Donald Ervin Knuth: *Computers & Typesetting. Vol. A: The TEXbook*. Addison-Wesley Publishing Company, Reading, Massachusetts. 1984.

[8] Korean TEX Society, `http://project.ktug.or.kr/ko.TeX`: *Korean TEX*. October 2007.

[9] Kuroki Yusuke: "Japanese TEX Environment for Cygwin". Slides for the 1st Asian TEX Conference. Kongju National University, South Korea. January 2008.

[10] Werner Lemberg: "The CJK Package: Multilingual Support beyond babel". TUG*boat*, Vol. 18, no. 3, pp. 214–224. September 1997.

[11] Werner Lemberg: "Unicode Support for the CJK Package". Hand-out version for the 1st Asian TEX conference. Kongju National University, South Korea. January 2008.

[12] Frank Mittelbach and Michel Goossens, with Johannes Braams, David Carlisle, Chris A. Rowley, Christine Detig and Joachim Schrod: *The LATEX Companion*. 2nd edition. Addison-Wesley Publishing Company, Reading, Massachusetts. August 2004.

[13] Okumura Haruhiko: "Japanese TEX. Past, Present, and Future". Slides for the 1st Asian TEX Conference. Kongju National University, South Korea. January 2008.

[14] Oren Patashnik: *BibTEXing*. February 1988. Part of the BibTEX distribution.

[15] John Plaice and Yannis Haralambous: "The latest developments in Ω". TUG*boat*, Vol. 17, no. 2, pp. 181–183. June 1996.

[16] Bernd Raichle: *Die Makropakete „german" und „ngerman" für LATEX 2ε, LATEX 2.09, Plain-TEX and andere darauf Basierende Formate*. Version 2.5. Juli 1998. Im Software LATEX.

[17] Tanaka Takuji: "Short Introduction to upTEX — Unicode pTEX". Slides for the 1st Asian TEX Conference. Kongju National University, South Korea. January 2008.

[18] Thành Hàn Thế: "Typesetting Vietnamese with VnTEX (and with the TEX Gyre fonts too!)". TUG*boat*, Vol. 29, no. 1, pp. 95–100. EuroBachoTEX 2007 proceedings. 2007.

[19] The Unicode Consortium: *The Unicode Standard Version 5.0*. Addison-Wesley. November 2006.